

# UNCLASSIFIED

AD NUMBER
ADB801471
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to DoD only; Administrative/Operational Use; Oct 1950. Other requests shall be referred through USAF School of Aviation Medicine, Randolph AFB, TX.
AUTHORITY
DTIC Form 55 dtd 30 Nov 2001, AF Institute for Environment Safety & Occupational Health Risk Analysis

THIS PAGE IS UNCLASSIFIED

Reproduced by



CENTRAL AIR DOCUMENTS OFFICE

WRIGHT-PATTERSON AIR FORCE BASE - DAYTON, OHIO

REEL-C

47471

A.T.I

993821

"NOTICE: When Government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the U.S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto."

UNCLASSIFIED

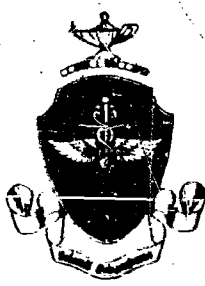
ATTN. 29 387

FILE COPY

UNITED STATES  
AIR FORCE

# *School of* **AVIATION MEDICINE**

DISCRIMINATORY ANALYSIS  
III. DISCRIMINATION OF ACCIDENT PRONE INDIVIDUAL



PROJECT NUMBER 21-49-004  
REPORT NUMBER 3  
(Formerly Project Number 21-02-105)

## **PROJECT REPORT**

# DISCRIMINATORY ANALYSIS

## III. DISCRIMINATORY OF ACCIDENT PRONE INDIVIDUALS

Harry M. Hughes, Ph.D.  
Research Associate  
Statistical Laboratory  
University of California, Berkeley

PROJECT NUMBER 21-49-004  
REPORT NUMBER 3\*

(Formerly Project Number 21-02-105)

\* A University of California report under  
Contract No. AF41(128)-8

USAF SCHOOL OF AVIATION MEDICINE

RANDOLPH FIELD, TEXAS

OCTOBER 1950

**Precis**

**OBJECT:**

To present an example of non-normal discriminatory analysis.

**SUMMARY:**

The mathematical statistical problem of the discrimination of accident prone individuals is presented with many illustrative examples from industrial and Air Force literature. The arguments in the literature for proneness are criticized and the need for better mathematical models of proneness is pointed out.

UNCLASSIFIED

ATI 99 382

(COPIES OBTAINABLE FROM CADO)

UNIVERSITY OF CALIFORNIA, STATISTICAL LAB., BERKELEY

DISCRIMINATORY ANALYSIS - III - DISCRIMINATORY OF  
ACCIDENT PRONE INDIVIDUALS

HARRY M. HUGHES OCT '50 43PP TABLES, GRAPHS

USAF SCHOOL OF AVIATION MEDICINE, RANDOLPH AIR FORCE  
BASE, TEX., USAF CONTR. NO. AF41(128)-8 (REPORT NO. 3)

AVIATION MED. CINE (19) BIOMETRICS  
GENERAL (0) STATISTICAL ANALYSIS

UNCLASSIFIED

## SECTION 1

### Introduction

In most of the problems treated in the foregoing\* the observable variables have in practice been assumed to have a normal distribution. But when one observes the number of accidents per person occurring in a reasonable period to a sample of persons, the distribution is not even approximately normal. Furthermore if one considers attached to each person a number representing his average accident rate--that is, the average number of accidents per year, say, over his adult lifetime--then the distribution of these rates in a certain population would certainly not be normal because of its non-negative character; and in practice this distribution appears to be generally too skew for normal approximation. However, it is obvious that these rates can never themselves be observed, so that our information as to their distribution must be inferred. For these reasons the study of accidents provides an excellent example of non-normal discriminatory analysis.

---

\* Report Number One, "Survey of Discriminatory Analysis", by Dr. J. Hodges; and Report Number Two, "Factor Analysis and Discrimination", by Dr. E. Fix.

But what is the discriminatory problem? First it must be established that different people under similar conditions have different accident rates, and hence that the problem exists of discriminating the high-rate persons from the low-rate. To solve the problem efficiently it is necessary to analyze the nature of the distribution of the rate and of associated variables. An implicit assumption has been made that such rate is constant over the years for a given person under a certain set of exposure conditions. The present part describes the attempts that have been made to solve these problems and to set up discriminant procedures, and indicates various points, both theoretical and experimental, which appear fruitful for further research.

## SECTION 2

### Existence of Individual Differences in Frequency of Accidents

The first efforts at reduction of accidents consisted of safety measures affecting alike all workers in a department or all citizens of a town, or all who passed a certain intersection. But it was early noted--industrially first, since individual records were there available--that different departments or types of work had average rates consistent with themselves but consistently different from those of other departments. This was to be expected, due to different conditions of exposure to risk. Yet even within a department, where all workers did the same thing for the same number of hours per week, it was noticed that "the bulk of the accidents occurred to a limited number of individuals."

It is appropriate to note at this point that the quoted statement, although it led the original workers into investigations which eventually substantiated by other means the existence of individual differences, is not itself such a substantiation. As Mintz and Blum (1949) adequately demonstrate, even if all persons have the same proneness, a large proportion of the accidents will occur to a small number of the individuals.



All investigators to date have started with the assumption that the chance of occurrence of an accident does not depend on the severity of the accident. That is, if a person has a high rate of scratches he will have a higher probability of having a serious accident than the person who has a low rate of scratches. On the basis of this assumption one is enabled to forecast or avoid disastrous consequences on the basis of comparatively harmless observations. In the event the assumption is wrong--and one might well argue that some persons are more careful in the face of known extreme danger than others--the studies still are valid for minor to moderate injuries and useful as they stand. The practical reason for choosing to study minor over major accidents is obvious; the statistical reason is that individual differences do not show up well when the average rate is low. If the average rate--even if it be minor accidents--is on the order of 0.3, say, then there is little room for variability in individual rates when we remember that the number of accidents is 0, 1, 2, ... and the length of observation is necessarily restricted. Thus it is desirable in making studies to find a department in which the accident rate is high and to use a unit period as long as possible subject to the other requirements of the analysis. With an average departmental accident rate of 4.0, say, there is much more room for individual variations in rate to appear if they exist, even in one observational period.

There have been attempts to establish statistically

the connection between major and minor accidents. One practical difficulty is that the figures for minor accidents represent only those which are reported whereas major accidents are faithfully recorded. In gathering data one must try to reduce the unreported accidents by some device, pressure, or incentive; and in interpreting data one must realize that it is really a tendency to have and to report accidents which is being measured. With these facts well in mind, Farmer and Chambers (1926) analysed six different trades in a group of 14,524 dockyard workers for one year with the results shown in Table 1. A major accident was defined as an accident involving one day or more lost time, and the correlation coefficient found between the number of major accidents and the number of minor accidents. Using an approximate formula for the probable error of the correlation coefficient when the

Table 1

Number in group	Trade	Major accidents Average rate	Minor Accidents Average rate	Correlation Coefficient
6,050	Laborers	.053	.534	.14
3,507	Shipwrights	.029	.535	.17
429	Boys	.051	.935	.25
295	Riveters	.078	.820	.25
181	Caulkers	.127	2.226	.31
639	Boilermakers	.052	.512	.36
2,634	Fitters	.029	.446	.06
339	Drillers	.056	.876	-.04
196	Smiths	.071	.806	-.02
54	Welders	.074	1.018	-.00

observations are normally distributed, Farmer and Chambers concluded that the first seven correlation coefficients, being greater than 2-1/2 times their estimated probable er-

rors, were significantly different from zero and hence indicated a positive correlation between major and minor accident rates. However these two variables are far from being normally distributed; in fact they take on only small integral non-negative values with a large number of zeroes. The regression may be far from linear, so that the correlation coefficient does not give a good description of the relationship; even if the regression is linear, the significance of the correlation coefficient depends on the relative sizes of the two average rates. For a discussion of the low values of  $r$  to be expected here due solely to the fact that we observe numbers of accidents rather than individual rates, see the early part of section 4. But one qualitative fact is obvious; the sample correlation coefficients lean heavily to the positive side of zero.

Having noted the cautions to be observed, consider now the question: do the accident rates of different persons under similar conditions differ significantly? A particular person is exposed to the risk of an accident for many instants of time with a small probability that the accident will occur at any particular instant. If that probability is constant then the number of accidents occurring in a particular period will follow the Poisson law:

$$(1) \quad P\{X=k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

where  $P\{X=k\}$  denotes the probability that the number of

accidents in a period will be  $k$ , and  $\lambda$  denotes the mean number of accidents per period for that particular person; i.e.,  $\lambda$  is that number attached to each person which has been defined as his accident rate under the given conditions and is to be distinguished from the actual number  $k$  of accidents befalling him in a period. Unfortunately it is not possible accurately to measure the  $\lambda$ 's of two people and hence to prove conclusively that they are different. It is theoretically possible to observe the actual accidents occurring in many periods, to estimate the two  $\lambda$ 's, and by normal approximation to test for a significant difference. but this requires a sufficient number of periods--a lifetime, perhaps--to justify the normal approximation to the mean of the Poisson. (It is implicitly assumed here that the rate does not change as time goes on. For a discussion of this assumption see section 4.) Instead one may observe many people under similar conditions of exposure over as long a time as is practicable, obtaining a distribution of the number of accidents occurring; i.e., the proportion of persons with 0 accidents, with 1 accident, with 2 accidents, etc.

If all the persons in such a group have the same rate  $\lambda$ , then the observations form a sample from a Poisson distribution. It is thus possible to test the hypothesis that all people under similar conditions of exposure have the same rate  $\lambda$  by testing the statistical hypothesis that the observations come from a Poisson distribution. Greenwood and Woods (1919) did this for 14 groups of women, from 50 to 750

in size, each in a particular type of work in munitions production. The best estimate of the fitted  $\lambda$  was of course the average observed rate; the fits were all so poor that not even a  $\chi^2$  test was considered necessary in order to reject the hypothesis that the observed numbers of accidents were from a Poisson distribution--and hence to reject the hypothesis that all persons in each group had the same rate. Newbold (1926) analysed 39 groups, from 22 to 440 in size, in various industrial operations by superimposing, on the observed frequency curves, fitted curves based on the average rate and fitted curves based on the number of persons having zero accidents. In both cases it was "very clear from these graphs that...groups are not homogeneous as regards accident risk."

For more precise indication of the lack of homogeneity, Newbold suggested the use of two criteria--the coefficient of variation and the relation of the average rate to the number of persons having zero accidents. The coefficient of variation, being the standard deviation divided by the mean, is  $\lambda^{-\frac{1}{2}}$  for a Poisson distribution; thus the observed coefficient of variation may be compared with the reciprocal root of the average accident rate, any divergence of the two indicating departure from Poisson. Lacking knowledge of the sampling distribution of the coefficient of variation, one has no reasonable significance criterion for saying when an observed coefficient of variation is too far away from the theoretical one. The same objection holds for

the other criterion suggested: Newbold presents a table of  $y = \log_e x$  so that entering the table with the observed proportion  $x$  who had zero accidents one may read off the average accident rate  $y$  to be expected from a Poisson distribution. Again the observed average rate and the theoretical rate diverge--but if one lays down a rule as to when the divergence is sufficiently great to deny equal tendencies to accident, how often is the decision wrong in repeated applications of the rule? Here some knowledge is needed regarding the joint distribution of zero frequency and mean in sampling--from a Poisson distribution at least.

However the data presented by Farmer and Chambers (1939) and by Newbold (1926), when fitted with a Poisson distribution, gives values of  $\chi^2$  corresponding to probabilities of less than .0001. This establishes statistically the existence of individual differences in accident rates under similar conditions of exposure and environment and hence of a quality called accident proneness. The next problem is to identify as early as possible the accident-prone; i.e., those with accident rates high in comparison with their fellow workmen. It is obviously futile to base such a classification on many years of observation if it is at all possible to make the classification sooner and thus reduce accidents. It is also futile, for statistical reasons, to attempt to make the classification on the basis of the number of accidents occurring to that person in a very short time since the random or sampling variations would mask the actual rates.

For example, consider an oversimplified situation in which there are 100 "good" persons with  $\lambda = 0.5$  accidents per year and 100 "poor" persons with  $\lambda = 1.5$  accidents per year all doing the same type of work. An attempt is made to eliminate, on the basis of the number of accidents incurred in one year, as many "poor" persons as possible. Tentatively consider the elimination of all who incur one or more accidents; by this rule we may expect to eliminate about 78 of the 100 "poor" ones. But at the same time 39 of the "good" persons would be eliminated. If the elimination were to take place at the end of two years instead of one year, those being dropped who incur two or more accidents in the two years, then 80 of the 100 "poor" could be eliminated at the expense of losing only 26 of the 100 "good." The improvement is not spectacular but indicates the danger in using a short time of observation. If the number of "poor" were originally smaller, say 40, the results would be even less satisfactory since 32 "poor" and 26 "good" would be eliminated in one year. The numbers computed here represent the probabilities of eliminating a person as a function of his accident rate  $\lambda$ ; they constitute the performance characteristic of the elimination rule, by which it is judged in comparison to other rules. To compute this performance characteristic--and hence to make the comparative judgment--it is necessary to know something of the distribution of the  $\lambda$ 's among the people being observed, or at least among the population from which the people being observed are drawn. Therefore section 3 will

contain an investigation of the  $\Lambda$  distribution.

An alternative approach to the identification of the accident prone is to discover some other measurements which may be made on a person, such as psychological or physiological tests, which alone or in combination would be sufficiently accurate predictors of the person's accident rate. The scores on such tests might well be normally distributed but the decisive variable is not. This approach, treated in section five, also requires knowledge of the distribution of accident proneness.





### SECTION 3

#### Distribution of Accident Proneness

The first attempt to explain the individual differences was made by Greenwood and Woods (1919) as follows: in a particular environment the original accident rate is the same for all persons, say  $\lambda_0$ . (Note that "accident rate" still refers to the long-range mean number  $\lambda$  of accidents per period.) As soon as a person experiences an accident then his rate becomes  $\lambda_1$ , regardless of who the person is. As soon as a person has had two accidents his rate becomes  $\lambda_2$ --and so on. Nine out of fourteen groups, fitted with the so-called "biased distribution" in which  $\lambda_0 \neq \lambda_1 = \lambda_2 = \dots$ , resulted in a  $\chi^2$  corresponding to a probability higher than .10--that is, were not significantly different in distribution at the 10 percent level. However, Greenwood and Woods preferred another explanation (see below) which sometimes had a smaller value of  $\chi^2$  (better fit) and had twelve out of the fourteen groups fit acceptably at the 10 percent level.

There was another reason given for the rejection of the bias explanation. For each of four groups of from 21 to 36 women the sample correlation coefficient was obtained between the number of accidents occurring in one period of three months and the number of accidents occurring to the

same person in a subsequent period of three months. The correlation coefficient was then computed for each of the four subgroups formed of women who had at least one accident in the first period. The investigators argued as follows: The persons removed had the same rate to start the second period as the first so their number of accidents in the two periods tends to be uncorrelated. Then their removal would undilute the original results so as to give higher correlation coefficients for the subgroups. However the latter compute to be lower; therefore the bias assumption is wrong. The reasoning is hardly precise and the data show little decrease in the correlation coefficient as a result of the removal of persons with no accidents in the first period; the groups are small, the observations decidedly non-normal, the probable errors apparently have been estimated assuming normality; all in all the evidence presented against the "biased distribution" is not convincing.

And yet the latter is the simplest form of a very logical structure. Greenwood and Yule (1920) gave the formulas for fitting a distribution with as many as five different values of  $\lambda$  instead of just two. However the fitting is fairly laborious and the accuracy of the estimates of the  $\lambda$ 's is unknown. Greenwood and Yule tried to find a simple flexible functional relation between  $\lambda_k$  and  $k$  but reported no success. Irwin, in a discussion of a paper by Chambers and Yule (1941), shows that if  $\lambda_k = \lambda_0 + ck$  with

$c > 0$  then the distribution of accidents will be a negative binomial, which will be seen in what follows to fit the observed data quite well. Thus we are led to consider this explanation as yet not sufficiently contradicted. However Irwin points out that the necessary restriction  $c > 0$  implies an increase in the average rate in time whereas the data show if anything a decrease in the average rate (see comments in section four). It still remains an open question whether or not there exists a simple functional relation between  $\lambda_k$  and  $k$  that would be comparatively simple to fit, would fit the data acceptably by  $\chi^2$  test, and would perhaps infer a reduction of average rate with experience.

The other principal theory of accident proneness is the one which assumes each person to maintain a constant  $\lambda$  under given environmental conditions with some distribution of this  $\lambda$  in the population. The first distribution to be tried was the normal--obviously a negative  $\lambda$  is meaningless but the actual  $\lambda$  distribution may be essentially normal nevertheless if the variance is sufficiently small with respect to the mean. Greenwood and Yule (1920) found the equations for the first few frequencies of the accident distribution, using a normal  $\lambda$  distribution. These equations indicate that in order to avoid fitted accident frequencies less than zero or greater than unity the average accident rate must exceed one-half the variance of the accident distribution. But that condition is not satisfied by most ob-

servations. Hence the normal distribution is of no use here even as an approximation.

The next likely distribution--one having a range from zero to infinity with both flexibility and a minimum of parameters for ease in fitting--was the Pearson type III curve or Gamma distribution:

$$(2) \quad f_{\lambda} = \frac{b^a}{\Gamma(a)} e^{-b\lambda} \lambda^{a-1}, \quad 0 \leq \lambda < \infty$$

where  $f_{\lambda}$  denotes the frequency function or probability density function and  $a$  and  $b$  are arbitrary positive constants. Greenwood and Woods found this distribution to have another advantage: it results in a negative binomial distribution of accidents, whence the moments are easily calculated. If  $X$  stands for the number of accidents incurred by a person in a period then

$$(3) \quad P(X=k) = \left(\frac{b}{b+1}\right)^a \frac{\Gamma(a+k)}{\Gamma(a) k! (b+1)^k}$$

$$(4) \quad m = E(X) = a/b$$

$$(5) \quad \sigma_X^2 = E[(X-a/b)^2] = \frac{a(b+1)}{b^2} = m + \frac{a}{b^2}$$

where  $a > 0$  and  $b > 0$  are constants to be determined from the observed distribution of accidents. By substituting the observed mean and standard error in these two equations, moment estimates of  $a$  and  $b$  may be obtained from which an expected distribution may be calculated and tested against the observed by calculation of  $\chi^2$ . This has been done by

Greenwood and Woods (1919) for their 14 groups of women munition workers with only two observed distributions significantly different from the expected;--and those two, it is pointed out, are the large groups unsegregated by departments so that the "guarantee of equal exposure to risk is much slighter than in other cases." Newbold (1926) found out of eight groups that two were significantly different at the 5 percent level and another two significantly different at the 10 percent level. Even this was achieved through judicious grouping. However the fit was much better than for any other considered hypothesis.

In 1927 Newbold reported another approach to the distribution problem which emphasized the usefulness of the negative binomial distribution in this connection. Starting without any assumption as to the form of the  $\lambda$  distribution, she derived various relationships between the lower order moments of the  $\lambda$ -distribution and the lower order moments of the observed X-distribution of accidents:

$$(7) \quad E(\lambda) = E(X) = m, \text{ say.}$$

$$(8) \quad \sigma_{\lambda}^2 = \sigma_X^2 - m$$

$$(9) \quad \mu_{3,\lambda} = \mu_{3,X} - 3\sigma_X^2 + 2m$$

$$(10) \quad \mu_{4,\lambda} = \mu_{4,X} - 6\mu_{3,X} - 6\sigma_X^2 m + 11\sigma_X^2 + 3m^2 - 6m$$

$$(11) \quad \rho_{X\lambda}^2 = \frac{\sigma_{\lambda}^2}{\sigma_X^2} = 1 - \frac{m}{\sigma_X^2}$$

where  $\mu_{i,\lambda}$  is the  $i^{\text{th}}$  central moment of the variable  $\lambda$  and

$P_{X\lambda}$  is the correlation coefficient between the accident rate  $\lambda$  of a person and the number of accidents  $X$  incurred by that person in one period. Using the observed moments of the  $X$  distribution on the right-hand side of equations (7)-(10) to estimate the moments of the underlying  $\lambda$  distribution, Miss Newbold analyzed seven groups as to the best-fitting Pearson curve. Four were found to be of type IJ, two of type VI J and one was impossible ( $s_X^2 \leq m$ ). Since type III is the borderline case between IJ and VI J it here obtains further support as being a useful model in this situation.

In a small study of 59 trolley car motormen in 1948 Brown and Ghiselli report nine distributions of which, although they did not note the fact, three have a variance smaller than the mean, thereby putting them in the "impossible" area. The difficulty appears to be in the quite small average rate coupled with the small size sample, although one of the impossible groups has an average rate of 1.14.

The most recent suggestion regarding the distribution of  $\lambda$  is by Mintz and Blum (1949) to the effect that it may consist of a continuous portion such as Pearson type III upon which is superimposed a small group of particularly prone individuals with  $\lambda = \lambda^*$ , a comparatively high value. Two of Newbold's distributions were thus fitted but no firm conclusions were drawn.

#### SECTION 4

##### Stability of Distribution of Accident Proneness

There are two types of stability to be investigated. One is the stability in time which has already been mentioned; i.e., does the person who has a high accident rate in one set of periods still have a high rate in a later set of periods? Note that "rate" still is defined to mean the underlying parameter  $\lambda$ , presumed constant for a person in given environment during a period at least. Stability in time would imply that  $\lambda$  remains constant as long as the environment is constant. The other type of stability is stability in kind; i.e., does the person who has a high accident rate for one kind of accident also have a high rate for another kind of accident? Here we cannot expect the two  $\lambda$ 's to be the same since the average rate of an entire group or population will in general be different for different kinds of accidents; some kinds will occur for everyone more often than other kinds. Creased fenders may appear more frequently in taxicab operations than flat tires. But stability in kind would imply that if Mr. Smashum has a higher  $\lambda$  for creased fenders than does Mr. Goezy then Mr. Smashum will also have a higher  $\lambda$  for flat tires than Mr. Goezy.



Consider first the verification of stability in time. The first technique, by Greenwood and Woods in 1919, was to correlate the number of accidents  $X_1$  and  $X_2$  occurring to each person in successive periods. These first computations were admittedly made on very small groups; the resulting four correlation coefficients ran from .37 to .72 with probable errors estimated at about .10. Newbold (1926) obtained a range of .20 to .57 in nine out of eleven groups with -.01 and +.05 in the other two groups. Calculated probable errors ran less than .10 but, as previously noted, the variables here being correlated are far from normal so that the sampling variation of the correlation coefficient may well be quite different. The consistently positive results offer qualitative support to the stability; however, one could wish for more precise knowledge of the significance of these figures.

In 1927 Miss Newbold presented, along with the moment computations displayed above, a formula for the correlation coefficient to be expected between the number of accidents  $X_1$  in one period and the number of accidents  $X_2$  in another period, assuming that each person will retain his  $\lambda$  through both periods, but not assuming anything about the distribution of the  $\lambda$ 's among the people:

$$(12) \quad \rho_{X_1 X_2} = \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{\sigma_{\lambda}^2 - m}{\sigma_{\lambda}^2}$$

where as before  $m = E(X) = E(\lambda)$ . This may be estimated by

inserting the mean and standard error of the observed accident distribution on the right:

$$(13) \quad r_e = \frac{s_x^2 - \bar{X}}{s_x^2}$$

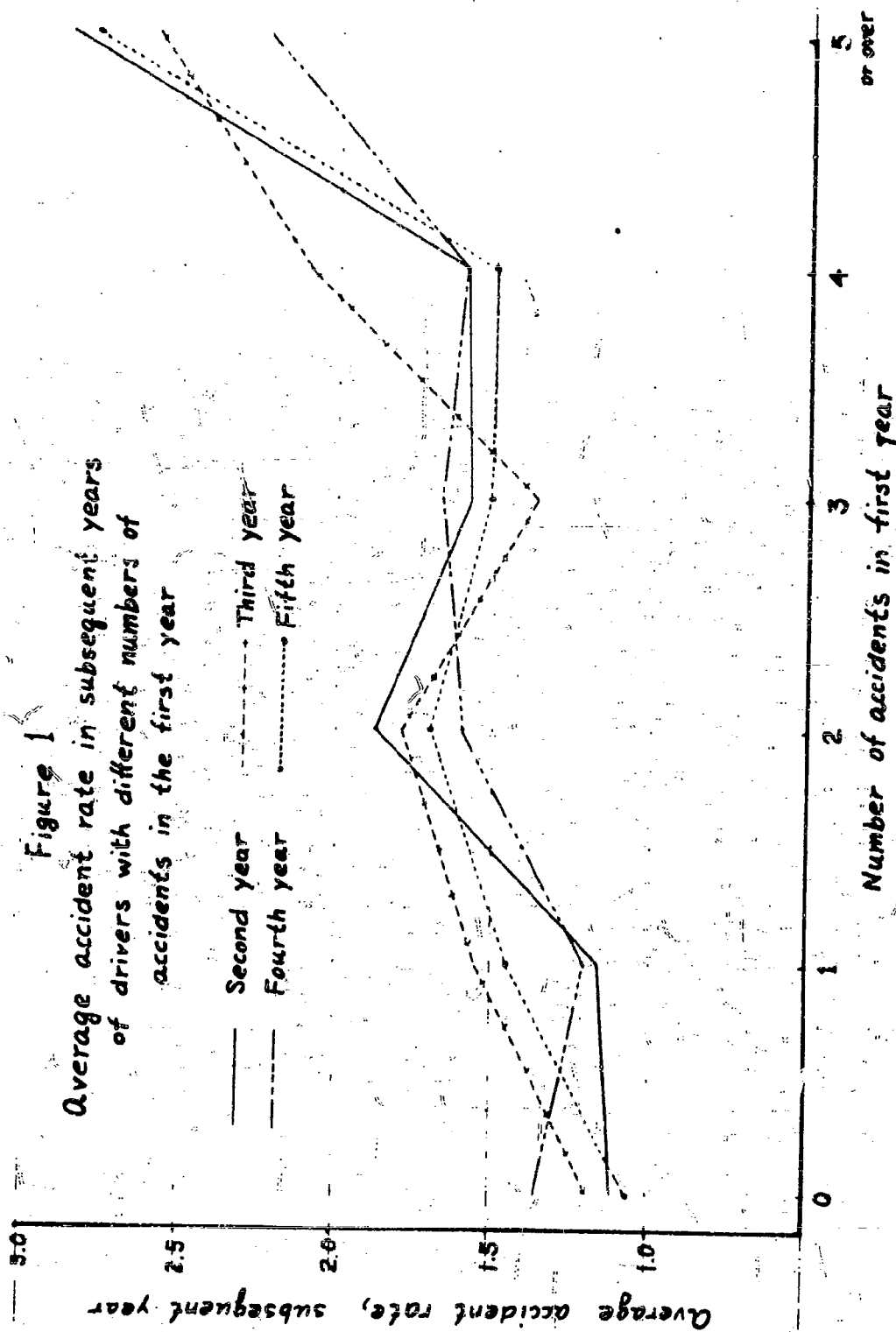
where

$$(14) \quad s_x^2 = \frac{1}{n-1} \sum (X - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n} \sum X,$$

the summations being taken over the sample. If both periods are combined for this estimation,  $n$  will be twice the number  $N$  of people observed. But nothing is known of the distribution of this statistic (13) or of the sample correlation coefficient

$$(15) \quad r_{x_1 x_2} = \frac{\sum x_1 x_2 - N \bar{x}_1 \bar{x}_2}{(n-1) s_{x_1} s_{x_2}}$$

It is not even known that either has the expectation  $\rho_{x_1 x_2}$  or that they have the same expectation. The important point is that the assumption of stability in time does not imply a correlation of unity between observed accidents in successive periods; with the two exceptions previously discovered, nine out of eleven of Newbold's sample correlations fell reasonably close to the corresponding values of  $r_e$ . Farmer and Chambers (1939) in a study of motor drivers found consistently positive values of  $r_{x_1 x_2}$  that again were very close to the corresponding values  $r_e$ , although they did not calculate the latter. It would be comforting to know something of the sampling variations in these statistics and



hence the significance level of these results; or perhaps a different approach to the verification of time stability would be more convincing.

A different approach was tried by Farmer and Chambers in 1959. Considering the number of accidents incurred in the first period to be indicative of the relative size of  $\lambda$ , they computed the average accident rate in the second period for the subgroup who had no accidents in the first period, the average rate for the subgroup who had one accident in the first period, the average rate for the subgroup having two accidents in the first period, etc. A graphic presentation of their results is given herewith in Figure 1, based on five one-year records of 166 bus drivers. These regressions show a gentle upward trend accentuated by the magnified vertical scale. But is the slope significantly different from zero? Or is it significantly different from what is to be expected, considering that the number of first-year accidents is at best a rough indication of  $\lambda$ ? These questions have not yet been answered, although they do not appear impossible of solution. This subject bears directly on the problem of discrimination--if the slope is not very steep then elimination of those with a large number of accidents in a trial period would not give efficient results. Farmer and Chambers were content to apply a t-test to the average rates of 0 group vs. 5-or-over group for each subsequent year. This showed a significant difference but is at best a very rough check of stability in time.

Stability in time, in the sense that  $\lambda$  remains constant for a given person and a given environment, implies that the average rate of a group will remain constant (within sampling variation) in successive periods of observation. Chambers (1941) stated that it was "quite clear from an inspection" of the average rates in five successive years of observations on 166 bus-drivers that the average rate diminishes with time:

	Year	1	2	3	4	5
(16)	$\bar{X}$	1.81	1.54	1.61	1.51	1.54
	$s_X$	1.7	1.4	1.5	1.3	1.4

Yet we do not know the sampling distribution of  $\bar{X}$ ; and if we use the normal approximation toward which it tends with large size samples, where  $\sigma_{\bar{X}}$  is on the order of .11, there is no significant contradiction to an assumption of five independent samples from the same  $X$ -distribution, and hence of a  $\lambda$ -distribution constant in time. This is particularly true if the first year is omitted under the assumption that the drivers were not trained until the beginning of the second year. The other data cited by Chambers, on 101 shipwrights, does however show a significant downward trend in the average rate, even with the omission of the first year. Consequently more data needs to be compiled and a more powerful method of analysis (such as the small-sample distribution of  $\bar{X}$ , perhaps) are both needed to establish or disprove a decrease in average rate in successive periods of observing a given group.

Even if such a trend is established, however, it does not contradict the time-stability of accident proneness. The  $\Lambda$ 's of each person may change in time, but if they change in such a way as to maintain their relative order on the  $\Lambda$  scale--that is, if Mr. A has a higher  $\Lambda$  than Mr. B in all periods--then there is still a stability of accident proneness, the individual differences. However, in order to have efficient discrimination it will be necessary first to establish the nature of this stability--how do the individual

$\Lambda$ 's change in time; do they maintain constant ratios; what is the relationship between the distribution of accidents in two successive periods?

In 1926 Farmer and Chambers introduced a device designed to separate the individual effect from the group effect and hence to permit studies of accident proneness even when the group average rates differed. This device was the "percentage accident rate," the observed number of accidents for a person in a given period divided by the average rate of the group for that period and expressed as a percentage. No attempt was made to establish the stability of this measure; neither is anything known of the sampling distribution of the PAR, of the mean of several PAR's, or of the correlation coefficient of the PAR with other variables--not even the degree of approximation to normality.

Consider now the existence of stability in kinds of accidents. The first investigation was in 1926 by Newbold, who found sample correlation coefficients of .56 and .32, with

estimated probable errors of .06 and .08, between accidents due to flying particles and other accidents, among 64 women in two periods of observation. Between factory and home accidents, separately computed for two groups of about 300 men each and two groups of about 150 women each, the sample correlation coefficients with estimated probable errors were .20  $\pm$  .04, .21  $\pm$  .03, .26  $\pm$  .05, and .31  $\pm$  .05. The driver accidents investigated by Farmer and Chambers in 1939 were classified into errors of judgment, over-runs, skids, blameless, and miscellaneous. The pairwise correlations of these various types of accidents ran from .08 to .39. It is interesting to note that the smaller correlation coefficients correspond to the kinds with the lowest rates. Again the question arises as to the significance of these results. Qualitatively they indicate some connection between the two types of accidents, but how good are the estimates of probable error, what is the highest correlation possible, what is the expected correlation? More generally, is there not some better way of delineating or testing the connection between two types of accidents than the correlation coefficient?

Farmer and Chambers used two other methods to check the stability in kind--a fourfold table and a regression. The fourfold table was made up by dividing the whole group into two broad subgroups according as their number of accidents of a particular kind was above or below the group average for that kind of accident; then  $\chi^2$  was computed for this distribution vs. a uniform distribution in the four cells. This

gave significant association in five out of twenty cases; but this test tends toward false significance since even if there is no association the table would not be expected to show a uniform distribution because of the skew and discrete character of the accident-distribution. The regression approach was similar to their regression attack on time-stability. The average rate of skids was calculated for the subgroup of drivers who had no errors of judgment, for the subgroup who had one error of judgment, for those who had two, etc. This series was repeated for each two kinds of accidents. A typical result is given in Table 2, showing some possibly U-shaped regression. But, what is to be expected? An analysis of regression based on an appropriate joint accident distribution would be helpful in judging the

Table 2

No. of Misc. Accidents	No. of Drivers	Average number of			
		errors	over-runs	skids	blameless
0	32	0.94	0.94	0.34	3.06
1	39	0.85	0.49	0.54	2.82
2	32	0.53	0.69	0.28	2.72
3	24	1.38	0.88	0.46	3.38
4	27	1.52	1.41	0.56	4.85
5-8	12	1.42	1.50	0.92	5.83

significance of this sample regression, and hence the existence of kind-stability. In 1948 a study of 59 trolley car motor-men and 34 motor coach operators by Brown and Ghiselli, with



accidents divided into five kinds, resulted in correlation coefficients from .22 down to .02, with one coefficient a negative .11. The authors' conclusions were that this data did not support stability in kind. However, we really do not know what constitutes a significantly positive coefficient in this situation, and the consistently positive result is hopeful. For example, since the average rate and variance of the observed distribution are approximately .43 and .45 respectively for collisions with trolley cars and for boarding accidents as well, the formula (13) may be used to find an approximation to the expected correlation coefficient:

$$(17) \quad r_e = \frac{s_x^2 - \bar{x}}{s_x^2} = \frac{.02}{.45} = .04$$

which is precisely the sample correlation coefficient that was observed. Despite the very low value then, the data do not really deny stability in kind. But is this just a coincidence? What about other low values? To compare two kinds of accidents with different average rates, say with the second rate equal to  $c$  times the first rate, it is possible to use a formula developed by Newbold (1927) for the correlation between the observed number of accidents in two periods of different length. It is possible, that is, provided

$$(18) \quad \sigma_{x_2}^2 - \bar{x}_2 = c^2(\sigma_{x_1}^2 - \bar{x}_1)$$

is roughly satisfied,  $\sigma_{x_1}^2$  being the variance of the distribution of the 1<sup>th</sup> kind of accident and  $\bar{x}_1$  being

the average rate for the  $i^{\text{th}}$  kind of accident. In such case

$$(19) \quad \rho_{X_1 X_2} = \frac{1 - \frac{\bar{X}_1}{\sigma_{X_1}^2}}{\sqrt{1 - \frac{\bar{X}_1}{\sigma_{X_1}^2} + \frac{1}{c} \frac{\bar{X}_1}{\sigma_{X_1}^2}}}$$

Considering then the data for collisions with motor vehicles as second kind to be correlated with either of the above-mentioned kinds,  $\bar{X}_2 = 12.22$  whence  $c = 28.4$ . Then

$$(20) \quad \rho_{X_1 X_2} = \frac{.0445}{\sqrt{.0445 + .0536}} = .16$$

as compared to the observed values  $r_{X_1 X_2} = .08$  and  $.22$ . It is obviously still necessary to know something of the sampling variability of  $r$ , even though  $\rho = .16$  removes the stigma from  $r = .22$ .

## SECTION 5

### Tests for Accident Proneness

To this point accident proneness has been considered only through its manifestation in the occurrence of accidents. In 1926 Farmer and Chambers gave a series of tests to a total of 651 people in an attempt to find some tests which would be useful in identifying those persons with relatively high accident rates--i.e., the accident-prone. Included were a reaction time test, a dotting accuracy test, a pursuit meter test, an ocular muscle balance test, and others; a priori considerations supported by significant inter-correlations led to a grouping into aestheto-kinetic coordination tests, temperamental instability tests and reasoning tests.

First a segregation of tests into those with discriminatory power and those without discriminatory power as regards accident-proneness was attempted by dividing the subjects into two subgroups according as they were above or below the mean test score and then computing the mean PAR (percentage accident rate; defined as the quotient: the observed number of accidents for a person over the average rate of the group) for each subgroup. The difference of these two means for a given test was then checked for significance

against its probable error, apparently calculated on the assumption of a normal distribution of the PAR's with equal variances.

Having selected the tests which were significant by this criterion, the next problem was to find a discriminant function of the scores in the significant tests--a single statistic whereby to rank the subjects in accident proneness. Two methods were suggested, although there is not the theoretical justification for them in this case that would be available if there were only a finite number of categories, say the accident-prone, the non-prone, and the medium prone, and if it could be determined in some preliminary sample to which of these categories each of the subjects belonged. Perhaps a corresponding development for this continuous category case can be worked out. At any rate, one of the methods suggested was a weighting of the individual standardized test scores by the differences in mean PAR computed as in the previous paragraph. The other method suggested was a weighting of the individual standardized test scores by regression coefficients of number of accidents on test score. To evaluate this procedure theoretically would require an analysis of regression in a joint distribution of a normal variate and a discrete variate, perhaps with a negative binomial marginal distribution.

To evaluate the weighting procedures experimentally, Farmer and Chambers divided the group of subjects into four subgroups by the quartiles of the weighted test score, and

found the mean PAR for each of the subgroups to be 62, 89, 112, and 139 percent. This was a disparity and trend sufficient to satisfy the authors that the weighted test score could successfully be used as a discriminant function. However it is not known what the probabilities of error are if the weighted test score is used; Farmer and Chambers themselves desired more data before drawing further conclusions. In 1929 the same analysis was reported for three years observation of 1042 Royal Air Force apprentices and 387 Royal Dockyard apprentices. It is not clear if PAR instead of number of accidents was used in this study although it seems probable. An encouraging fact in this study is that the weightings (regression coefficients) were found to be fairly constant from group to group. This would appear to indicate that the tests are stable indicators of accident proneness. However, the regression is of number of accidents on test score rather than of accident proneness on test score. For this reason one might expect the regression coefficients to vary with different environments; i.e., with different average accident rates. Perhaps a better measure would be regression coefficients of PAR on test score since this measurement tends to eliminate environmental effect. In any event, a theoretical study is needed of the meaning and invariance of the regression coefficients and correlation coefficients in such situations. In 1933 and again in 1939 Farmer and Chambers recognized this need, noting the lack of normality and the necessity of linearity of regression to make these methods useful.

## SECTION 6

### Discriminant Action

The discriminant problem connected with accident proneness is to separate those persons with a small  $\lambda$  from those with a large  $\lambda$  on the basis of several observations, such as the number of accidents in each of several periods or the scores on several tests. Such separation will be made on the basis of a rule, and it is ordinarily desired to obtain an optimum rule.

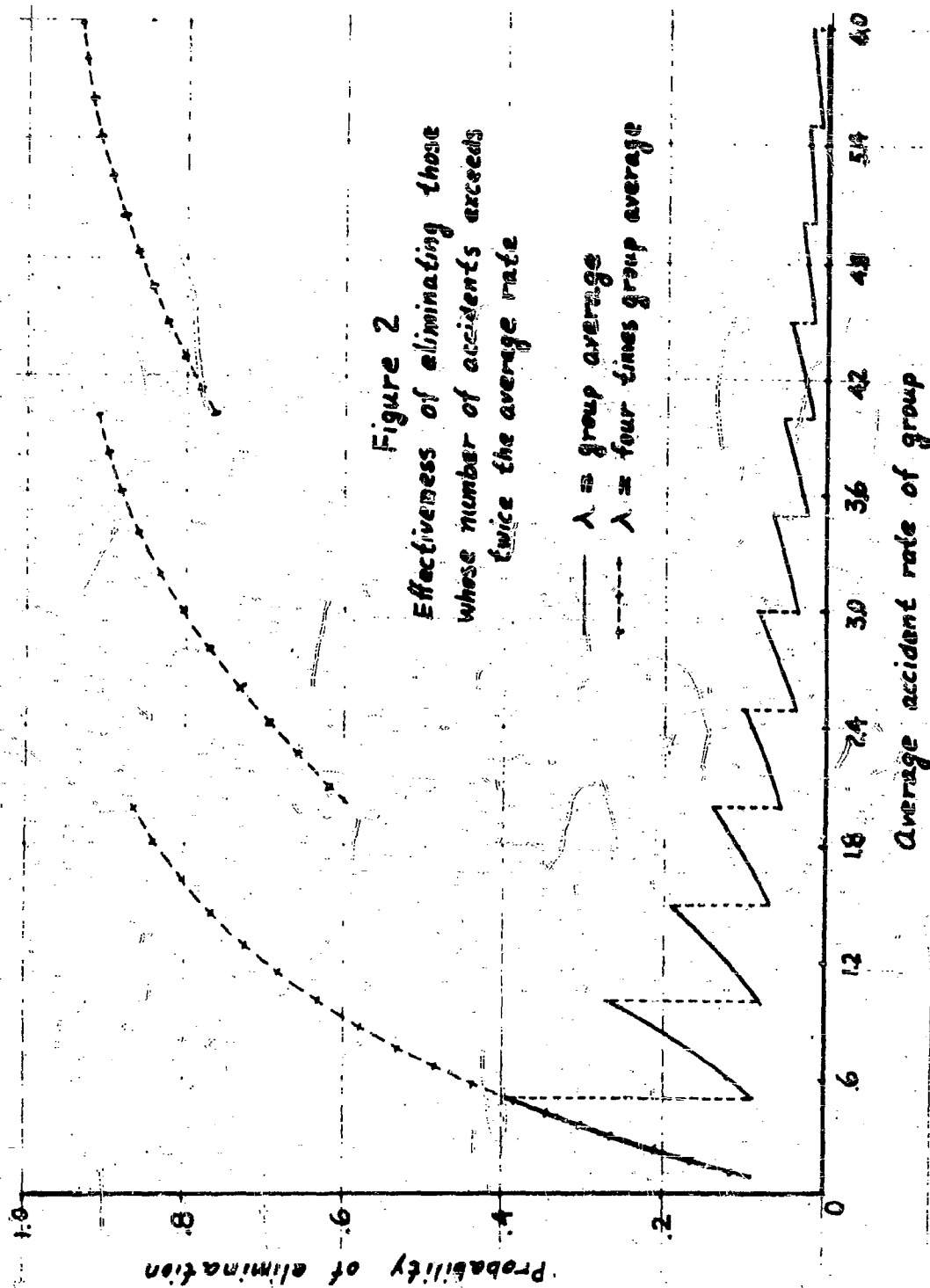
Despite the detailed regression, correlation, and  $\chi^2$  analyses of the principles underlying accident distributions, not much has resulted to date in dependable rules for discriminant action. The earliest attempt in this direction was a suggestion by Greenwood and Woods (1919) to eliminate those whose number of accidents exceeds twice the average rate. No analysis of the consequences of such a rule was given. The probability of being eliminated by such a rule is easily obtainable from the Poisson distribution, however, and is given in Table 3, for any particular individual.

Table 3  
Probability of Incurring More Accidents  
Than Twice the Average Rate

$\lambda$ Average rate	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0-.49	.593	.632	.777	.865	.918	.950	.970	.982	.993
.50-.99	.090	.264	.442	.594	.713	.801	.864	.908	.959
1.00-1.49	.014	.080	.191	.323	.456	.577	.679	.762	.875
1.50-1.99	.002	.019	.066	.143	.242	.353	.463	.567	.735
2.00-2.49	.000	.004	.019	.053	.109	.185	.275	.371	.560

Another way to judge the effectiveness of this plan is to examine the probability of eliminating the average person (one whose  $\lambda$  equals the group average rate) and the probability of eliminating a person whose  $\lambda$  is some multiple, say four times the average rate. Figure 2 exhibits these two graphs. An important aspect of both Table 3 and Figure 2 is the dependence of the effectiveness on the average accident rate of the group. It is particularly important to note that for small average rates the rule does not distinguish at all between the "average person" and the person whose rate is four times the average.

It has already been described how Farmer and Chambers proposed to eliminate one-fourth of the subjects--those below the third quartile in the weighted test scores. In 1959 they demonstrated the effect of various removal rules upon





the average accident rate of the group studied. By removing that 28 percent of the drivers who had three or more accidents the first year, the average rate in the succeeding four years would have been reduced to 93 percent of its actual value. By using both criteria, 44 percent would have been removed and the average rate reduced to 87 percent. This is a high cost for a small reduction, and would not be practical.

Herdan (1943) suggested the use of two-factor factor analysis to estimate  $\lambda$  of a person by a linear combination of his PAR for various kinds of accidents, the coefficients to be obtained by correlations in the usual manner of factor analysis--see part 2 of this report.

It would seem that the possibilities of obtaining an optimum rule have not been explored; in such direction there is a need of further investigation.

One trend in the work to date should be noted and discussed in this connection: the contention that since only correlations less than unity can occur between a rate  $\lambda$  and number  $X$  of accidents incurred, there is no point in attempting a discrimination. This arose from Newbold's (1927) formula

$$(21) \quad \rho_{X\lambda}^2 = \frac{\sigma_X^2 - m}{\sigma_X^2}$$

where  $m = E(X) = E(\lambda)$ . [Note parenthetically from (12) that  $\rho_{X\lambda}^2 = \rho_{X_1 X_2}$ ]. A clear exposition of this viewpoint

is given by Mintz and Blum (1949). Their approach is to examine that proportion of  $\sigma_X^2$  which is due to the regression on  $\lambda$  in an analysis of variance; however this is just  $\rho_{X\lambda}^2$ . The conclusion is that the regression will not account for as much as 50 percent of the variance in many cases. That is just another way of saying that even if we knew the  $\lambda$  of each person precisely, we would still be unable to forecast the number of accidents precisely. A correct conclusion, but what is the consequence on the discriminant problem? Presumably the discriminant problem has as its motive the reduction of the total number of accidents in a period by the location and subsequent elimination, transfer, and/or treatment [see Weinerman (1949)] of the high- $\lambda$  persons. The present question is: even if a perfect test for accident proneness is achieved, will the elimination of such persons reduce the total number of accidents significantly? One answer is that  $E(X) = E(\lambda)$  so that if the large  $\lambda$ 's are eliminated then the mean  $\lambda$  will be reduced and hence the mean number of accidents. How efficiently this can be done depends on the distribution of  $\lambda$  and is therefore a subject for further inquiry.

## BIBLIOGRAPHY

BLUM, M.L., (See Mintz, A.)

BRACKENMAN, E. E. (See Glocombe, C. S.)

BROWN, C. W., and Ghiselli, E. E., 1947. Factors related to the proficiency of motor coach operators. J. Appl. Psych., 31, 477-479.

1948. Accident proneness among street car motormen and motor coach operators. J. Appl. Psych., 32, 20-23.

CHAMBERS, E. G., and Yule, G. U., 1941. Theory and observation in the investigation of accident causation. Supr. J.R.S.S., 7, 89-109.

(See Farmer, E.)

COBB, P. W., 1940. The limit of usefulness of accident rate as a measure of accident proneness. J. Appl. Psych., 24, 154-159.

FARMER, E., and Chambers, E. G., 1926. A psychological study of individual differences in accident rates. I.F.R.B., 38.

1929. A study of the personal qualities in accident proneness and proficiency. I.H.R.B., 55.

1939. Study of accident proneness among motor drivers. I.H.R.B., 84.

and Kirk, F. J., 1933. Tests for accident proneness. I.H.R.B., 68.

GHISELLI, E. E. (See Brown, C. W.)

GREENWOOD, M., and Woods, H. M., 1919. The incidence of industrial accidents. I.F.R.B., 4.

and Yule, G. U., 1920. An inquiry into the nature of frequency distributions representative of multiple happenings. J.R.S.S., 83, 255.

HERDAN, G., 1943. The logical and analytical relationship between the theory of accidents and factor analysis. J.R.S.S., 106, 125-142.

I.H.R.B., 1942. The personal factor in accidents. I.H.R.B., Emergency Report No. 3.

KIRK, F. J. (See Farmer, E.)

MINTZ, A. and Blum, M.L., 1949. A re-examination of the accident proneness concept. J. Appl. Psych., 33, 195-211.

MUSCIO, B. (See Osborne, E.E.)

NEWBOLD, E. M., 1926. A contribution to the study of the human factor in the causation of accidents. I.F.R.B., 34.

1927. Practical applications of the statistics of repeated events. J.R.S.S., 90, 487-547.

OSBORNE, E. E., Vernon, H. M., and Muscio, B., 1922. Two contributions to the study of accident causation. I.F.R.B., 19.

SLOCOMBE, C. S. and Brakeman, E. E., 1930. Psychological tests and accident proneness. J. Psych., 26, 29-38.

U. S. BUREAU OF PUBLIC ROADS, 1938. Motor vehicle traffic conditions in the United States: The accident-prone driver. House Document No. 462, Part 6, 75th Congress, Third Session.

VERNON, H. M. (See Osborne, E. E.)

WEINERMAN, E. R., 1949. Accident-proneness: a critique. Amer. J. Publ. H., 39, 1527-1530.

WOODS, H. M. (See Greenwood, M.)

YULE, G. U., 1924. The function of statistical method in scientific investigation. I.F.R.B., 28.

(See Chambers, E. G.; Greenwood, M.)

Reproduced by



CENTRAL AIR DOCUMENTS OFFICE

WRIGHT-PATTERSON AIR FORCE BASE - DAYTON, OHIO

REEL-C

47471

A.T.I

99382

**"NOTICE:** When Government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the U.S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto."

UNCLASSIFIED

UNCLASSIFIED

ATI 99 382 *338400* (COPIES OBTAINABLE FROM CADO) *P1213*

UNIVERSITY OF CALIFORNIA, STATISTICAL LAB., BERKELEY

DISCRIMINATORY ANALYSIS - III - DISCRIMINATORY OF  
ACCIDENT PRONE INDIVIDUALS

HARRY M. HUGHES OCT'50 43PP TABLES, GRAPHS

USAF SCHOOL OF AVIATION MEDICINE, RANDOLPH AIR FORCE  
BASE, TEX., USAF CONTR. NO. AF41(128)-8 (REPORT NO. 3)

AVIATION MEDICINE (19) *12/16* *8* BIOMETRICS  
GENERAL (0) STATISTICAL ANALYSIS

UNCLASSIFIED

**AD-B801 471**



+